

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

RECEIVED

JAN 16 1995

OSTI

TITLE: BAYESIAN INFERENCE AND THE ANALYTIC CONTINUATION OF IMAGINARY-TIME QUANTUM MONTE CARLO DATA

Authors: James E. Gubernatis, T-11  
Janez Bonca, T-11  
Mark Jarrell, Univ. of Cincinnati

SUBMITTED TO: Proceedings of Max Ent. '95 Workshop; Editors: R. Silver and K. Hanson; Publisher: Kluwer Scientific

By acceptance of this article, the publisher recognized that the U S Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution or to allow others to do so for U S Gove National Laboratory requests that the publisher identify this article as work performed in the auspices of the U S Department of Energy.

Los Alamos Los Alamos National Laboratory  
Los Alamos, New Mexico 87545

FORM NO. 836 R4  
ST. NO. 2629 5/81

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

**DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

# BAYESIAN INFERENCE AND THE ANALYTIC CONTINUATION OF IMAGINARY-TIME QUANTUM MONTE CARLO DATA

J. E. GUBERNATIS AND J. BONČA  
*Theoretical Division, Los Alamos National Laboratory  
Los Alamos, NM 87545, U.S.A.*

AND

MARK JARRELL  
*Department of Physics, University of Cincinnati  
Cincinnati, OH 45221, U.S.A.*

**Abstract.** We present a brief description of how methods of Bayesian inference are used to obtain real frequency information by the analytic continuation of imaginary-time quantum Monte Carlo data. We present the procedure we used, which is due to R. K. Bryan, and summarize several bottleneck issues.

**Key words:** entropic prior, maximum entropy, analytic continuation

## 1. Introduction

Most often, discussions about the application of methods of Bayesian inference focus on the choice of prior probabilities. The choice of the likelihood function is usually viewed as less problematic. Typically, the natural choice is that of Gaussian distribution. In this paper, we will discuss a successful application of the methods of Bayesian inference to the problem of the analytic continuation imaginary-time data obtained from quantum Monte Carlo simulations. Here, a "natural" choice of the prior probability was the entropic prior. The choice of the likelihood function as a Gaussian distribution was an unnatural one but the only workable one.

The analytic continuation problem is a very ill-posed one. It amounts to numerically performing an inverse, two-sided Laplace transformation of noisy and incomplete data. Bayesian methods are used to develop a procedure to regularize the problem. A major difficulty in achieving success was overcoming the naturally non-Gaussingly-distributed data produced by the simulation. To achieve such a Gaussian distribution, large amounts of data were produced to force the central limit theorem to hold approximately. Experience shows that unless the data used as input to the analytic continuation problem consistent with the procedures used to perform the continuation, then unsatisfying, and often incorrect, results follow.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

While the need for such consistency in retrospect seems obvious, it was overlooked in several of the initial numerical approaches to the problem.

In what follows, we first will summarize and define the problem of the analytic continuation of imaginary-time quantum Monte Carlo data. Next, we will state the Bayesian approach we used. We simply borrowed a procedure proposed by Bryan [1]. In this context, we will give several illustrations showing why both the generation and statistical quality of the data can be bottleneck issues. We conclude with a summation. We comment that recently two of us wrote a review article that details our methods and procedures [2].

## 2. The Analytic Continuation Problem

For finite-temperature, quantum Monte Carlo simulations, the analytic continuation problem is “Given noisy and incomplete estimates for the imaginary-time correlation function  $G(\tau)$ , what is the spectral density  $A(\omega)$ ?” The imaginary-time correlation function (or many-body Green’s function) arises naturally in the field-theoretic approach to quantum many-body problems [3]. The analytic properties of this approach leads to the association of a spectral density to each type of many-body Green’s function. The spectral density often exhibits in its structure (its peaks and their location) information about the subtle microscopic processes taking place in physical systems. Often these functions can be measured experimentally. Theoretical physicists and chemists find the computation of these functions important and useful for the physics they reveal about the physical models being studied and about the relevance of the models to real physical systems.

Many-body quantum theory states [3] that  $G(\tau)$  and  $A(\omega)$  are related by

$$G(\tau) = \int_{-\infty}^{+\infty} d\omega \frac{e^{-\tau\omega} A(\omega)}{1 \pm e^{-\beta\omega}}$$

The  $\pm$  refers to Fermi and Bose statistics,  $\tau > 0$ , and  $\beta$  is the inverse of the temperature. A precise definition of  $A(\omega)$  can be given in terms of the eigenvalues and vectors of the Hamiltonian specifying the physical system, quantities which are very hard to obtain in part because the number of eigenstates generally increases exponentially with the size of the system. This situation points to utility for a Monte Carlo simulation of the problem. Depending on the definition of  $G(\tau)$ ,  $A(\omega)$  is related to such experimentally measurable quantities as the photoemission spectra, optical conductivity, dynamic magnetic susceptibility, etc. These quantities, not the real-time correlations, is usually all that is measured.

### 2.1. THE PROBLEM

The main difficulty in the analytic continuation problem is that at large positive and negative frequencies the kernel

$$K(\tau, \omega) = \frac{e^{-\tau\omega}}{1 \pm e^{-\beta\omega}}$$

is exponential. This condition makes  $G(\tau)$  insensitive to the high-frequency details of  $A(\omega)$ . In terms of obtaining  $A(\omega)$  from  $G(\tau)$ , this insensitivity leads to an ill-posed problem, and the ill-posedness implies an infinite-number of solutions exist. The task is to select from this infinity of solutions one that is "best" by some criterion.

## 2.2. GENERAL FEATURES

The spectral density satisfies

$$\begin{aligned} A(\omega) &\geq 0 \\ \int d\omega A(\omega) &< \infty \end{aligned}$$

These features allows us to interpret  $A(\omega)$  as a probability function. Often, the bound on the integral of the spectral density is precisely known or is computable by the simulation. In these cases, called *sum rules*, the bound also reveals physical information about the problem.

The correlation functions  $G(\tau)$  satisfy

$$G(\tau + \beta) = \pm G(\tau)$$

which allows us to restrict  $0 \leq \tau < \beta$ . Within this range

$$G(\beta - \tau) = G(\tau)$$

$G(\tau)$  is usually bounded from above and below.

## 2.3. IMAGINARY-TIME QUANTUM MONTE-CARLO

Quantum Monte Carlo simulations at finite temperatures are done in imaginary time  $\tau$ . The Wick rotation,  $it \rightarrow \tau$ , transforms Schroedinger's equation for a Hamiltonian operator  $H$

$$i \frac{\partial \psi}{\partial t} = H \psi$$

and its formal solution

$$\psi(t) = e^{-iHt} \psi(t=0)$$

into

$$\frac{\partial \psi}{\partial \tau} = -H \psi$$

and

$$\psi(\tau) = e^{-\tau H} \psi(\tau=0)$$

Other operators evolve and transform as

$$A(t) = e^{itH} A e^{-itH} \rightarrow A(\tau) = e^{\tau H} A e^{-\tau H}$$

The Wick rotation thus transforms oscillatory exponentials  $e^{-itH}$  into diffusive ones,  $e^{-\tau H}$ . The importance of such exponential lies with the dynamics of quantum mechanical systems.

Why perform the Wick rotation? From the point of view of performing a Monte Carlo simulation, at long times real-time such simulations are inefficient: sampling on smaller and smaller times scales becomes necessary as time increases just to achieve proper self-cancelations. Quantum simulations, however, are readily (and naturally) performed in imaginary-time.

### 3. A Solution Path

The functional form of the analytic continuation problem clearly reveals its ill-posed nature. In general, the solution of such a problem requires a regularizer. In the present situation, we have the added difficulties of the estimates of  $G(\tau)$  obtained from the simulations being incomplete and noisy. They are incomplete because we only determine  $G(\tau)$  at discrete values of  $\tau$ . They are noisy because of the nature of a Monte Carlo simulation.

To proceed, we first convert the integral equation to a linear system of equations

$$G_i = \sum_j K_{ij} A_j$$

where  $G_i = G(\tau_i)$ ,  $K_{ij} = K(\tau_i, \omega_j)$ , and  $A_i = A(\omega_i) \Delta\omega_i$ . Next, we pretend to consider a constrained least-squares approach, i.e., we consider seeking  $A_i$  that maximizes

$$Q = \alpha S - \frac{1}{2} \chi^2$$

$$\begin{aligned} \chi^2 &= \sum_{i,j} (\bar{G}_i - G_i) [C^{-1}]_{ij} (\bar{G}_j - G_j) \\ &= \sum_{i,j} (\bar{G}_i - \sum_k K_{ik} A_k) [C^{-1}]_{ij} (\bar{G}_j - \sum_k K_{jk} A_k) \end{aligned}$$

and

$$\bar{G}_i = \frac{1}{M} \sum_{m=1}^M \bar{G}_i^{(m)}$$

$C$  is the covariance matrix

$$C_{ik} = \frac{1}{M(M-1)} \sum_{m=1}^M (\bar{G}_i - \bar{G}_i^{(m)}) (\bar{G}_k - \bar{G}_k^{(m)})$$

The  $\bar{G}_i^{(m)}$  are statistically independent estimates of  $G_i$ ;  $S$  is the information theory entropy

$$S = \sum_i [A_i - m_i - A_i \ln(A_i/m_i)]$$

with  $m_i$  being the default model.

There are several things being assumed. One is the entropic prior. The rationale for this choice follows from the observation that the possible interpretation of the spectral density as a probability function and the utility of the principle of maximum entropy for assigning probabilities. In practice, the choice is very convenient because of the ease of insuring the non-negativity of the computed spectral density. Another assumption is the choice of a multivariate Gaussian for the likelihood function. We know of no other analytic form that fits or can be made to match the distribution of the data, short of mindless function fitting that most likely would be needed for each simulation.

What does one chose for  $\alpha$ ? We opt for the following [1]

$$\begin{aligned} \langle A \rangle &= \int \mathcal{D}A d\alpha A(\alpha) \Pr[A, \alpha | \bar{G}] \\ &= \int \mathcal{D}A d\alpha A(\alpha) \Pr[A | \bar{G}, \alpha] \Pr[\alpha | \bar{G}] \\ &\approx \int d\alpha \hat{A}_\alpha \Pr[\alpha | \bar{G}] \end{aligned}$$

In the above,

$$\begin{aligned} \Pr[A | \bar{G}, \alpha] &= \Pr[\bar{G} | A] \Pr[A | \alpha] \Pr[\alpha] \\ &= \frac{e^{-\frac{1}{2}x^2}}{Z_L} \frac{e^{\alpha s}}{Z_S(\alpha)} \Pr[\alpha] \\ &= \frac{e^Q}{Z_L Z_S(\alpha)} \Pr[\alpha] \end{aligned}$$

where  $Z_L$  and  $Z_S(\alpha)$  are the normalization constants for the likelihood function and the prior. Also,

$$\begin{aligned} \Pr[\alpha | \bar{G}] &= \int \mathcal{D}A \Pr[\bar{G} | A] \Pr[A | \alpha] \Pr[\alpha] \\ &= \Pr[\alpha] \int \mathcal{D}A \frac{e^Q}{Z_L Z_S} \end{aligned}$$

To proceed, for a fixed value of  $\alpha$ , we solve

$$\left. \frac{\delta Q}{\delta A} \right|_{A=\hat{A}_\alpha} = 0$$

to find  $\hat{A}_\alpha$ . (For a given value of  $\alpha$ ,  $Q$  is a convex function of  $A_i$ .) Then, we numerically perform

$$\bar{A} = \int d\alpha \hat{A}_\alpha \Pr[\alpha | \bar{G}]$$

The functional integration over  $A$  and the integral over  $\alpha$  are done numerically.

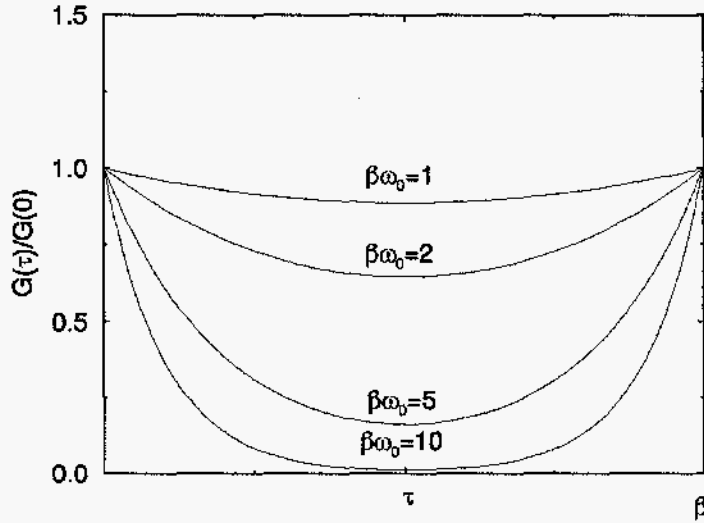


Figure 1. The exact Green's function as a function of  $\tau$  for a simple harmonic oscillator.  $\omega_0$  is the oscillator frequency and  $\beta$  is the inverse of the temperature.

#### 4. Bottlenecks

Only for a very large number of statistically independent measurements does the distribution of the Monte Carlo data approach the assumed multivariate Gaussian. The results are very sensitive to the extent to which the central limit theorem can "save the day." There several other issues regarding the data. Compared to Monte Carlo simulations where the analytic continuation is not an objective, the data in the continuation problem require a more careful statistical characterization, particularly with respect to statistical independence. (The assumed likelihood function assumes statistical independence.) Also the variance of the Monte Carlo results needs a greater reduction. (This need is a consequence of the ill-posedness of the problem.) The techniques we used to statistically characterize and qualify the data are discussed in [2].

In Fig. 2, we show the exact Green's function for a single quantum harmonic oscillator (or well). Those obtained from almost any simulation look like these. In general, some "characteristic energy scale"  $\omega_0$  exists. When it is smaller than the the temperature ( $\beta\omega_0 < 1$ ), the Green's function is flat. If one envisions error bars on the measured values, then a number of measurements are "within the error" of each other, and hence not all the measured data make independent contributions. When it is larger then the temperature ( $\beta\omega_0 > 1$ ), only in a small region near  $\tau = 0$  and by symmetry near  $\tau = \beta$  is the relative error small. Additionally, the smoothness of the curve implies that the different values of  $G(\tau)$  are correlated with one another. The correlation is the reason why the covariance matrix appears in the likelihood function.



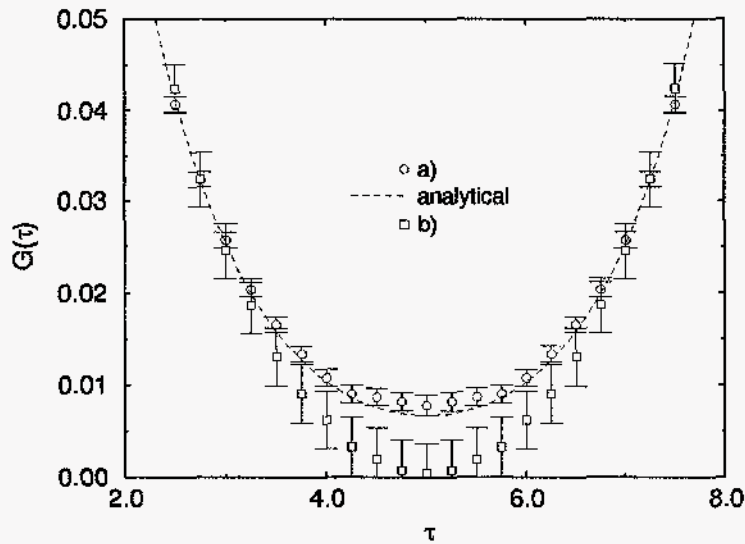


Figure 2. Near  $\tau = \beta/2$ , the Green's function (dashed line) from Fig. 2 for the  $\beta\omega_0 = 10$  case compared with the results obtained by two different quantum Monte Carlo algorithms. The results from the non-ergodic algorithm lie below the exact result.

In Fig. 3, we shown a portion of Fig. 2 (the dashed line) near  $\tau = \beta/2$  together with the values of the Green's functions computed for the same problem from two slightly different quantum Monte Carlo algorithms. The results which show the greatest deviation from the exact results were obtained with an algorithms that was slightly non-ergodic. We point out the reduced y-axis scale compared to Fig. 2. If we used in Fig. 3 the same scale as in Fig. 2, the differences between the results from the two simulations would be barely noticeable. Figure 4 illustrates that these small differences can make a big difference in the results [4]: the exact  $A(\omega)$  should be a  $\delta$ -function positioned at 1 with a weight (sum rule) of 0.5. The result from the ergodic algorithm gives a peak at the right location with the right weight. The width of the peak will narrow as the statistical error associated with data narrows (i.e., if more data is used). The result obtained with the other algorithm is broader and incorrectly placed and weighted. The breadth can be reduced by reducing the statistical error of the data, but this will not overcome the systematic error caused by the algorithm.

## 5. Remarks

The methods [1] briefly discussed provide a framework to approach the analytic continuation problem in which the assumptions and approximations in the approach can be clearly defined. They opened a completely new set of opportunities for the applications of quantum simulations. Several other important points are: (1) There is nothing about the methods that is intrinsic to quantum Monte Carlo

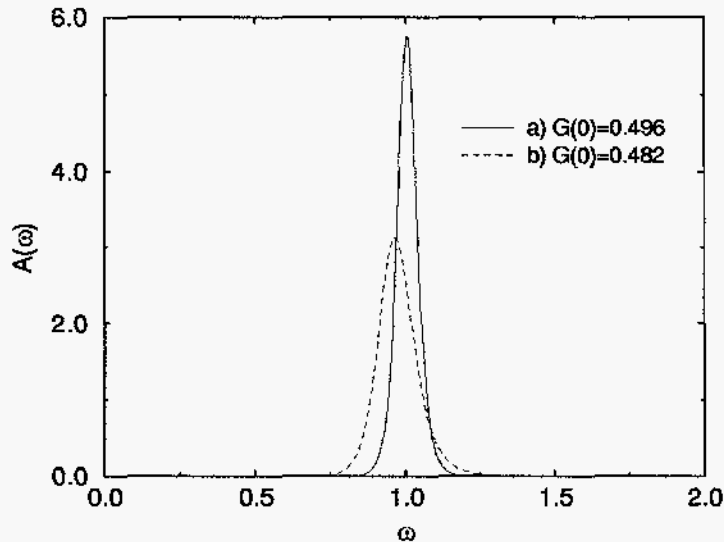


Figure 3. The spectral densities obtained from the simulation data partially represented in Fig. 3. The result (dashed line) obtained from the non-ergodic algorithm is improperly positioned and weighted. For the particular physical model, the value of  $G(0)$  gives the value of the sum rule, i.e., the expected value of the integral of  $A(\omega)$ .

or analytic continuation problems. (2) The Bayesian methods lead to a procedure with no adjustable parameters. This removes one of the difficulties often associated with other regularization methods. We also remark that the requirement for Gaussian-distributed data is a result of a particular choice of the likelihood function. Making the data as consistent as possible with this assumption is the hardest part of the problem and is the principal source of the increased computational cost associated with this problem. Recently, we have successfully obtained the actual real-time Green's function from the imaginary-time one [4].

This work was supported in part by the U. S. Department of Energy. We also gratefully acknowledge the influence and past collaborations with Devinder Sivia and Richard Silver.

## References

1. R. K. Bryan, "Maximum entropy analysis of oversampled data problems," *Eur. Biophys. J.*, **18**, pp. 165-174, 1990.
2. M. Jarrell and J. E. Gubernatis, "Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data," *Phys. Reports*, to appear.
3. See for example, A. Abrikosov, L. Gorkov, and I. Dzyaloshinskii, *Quantum Field-Theoretical Methods in Statistical Physics*, Pergamon, Oxford, 1965.
4. J. Bonča and J. E. Gubernatis, "Real-time dynamics from imaginary-time quantum Monte Carlo simulations: test on oscillator chains," *Phys. Rev E*, to appear.